Mon. Not. R. Astron. Soc. **361,** 887–896 (2005) doi:10.1111/j.1365-2966.2005.09213.x

The analysis of irregularly observed stochastic astronomical time-series – I. Basics of linear stochastic differential equations

Chris Koen[1,2,3]

[1]*Department of Statistics, University of the Western Cape, Private Bag X17, Bellville, 7535 Cape, South Africa*
[2]*Department of Mathematics & Statistics, University of Johannesburg, PO Box 524, 2006 Auckland Park, South Africa*
[3]*SAAO, PO Box 9, Observatory, 7935 Cape, South Africa*

Accepted 2005 May 19. Received 2005 April 7; in original form 2004 September 10

ABSTRACT

The theory of low-order linear stochastic differential equations is reviewed. Solutions to these equations give the continuous time analogues of discrete time autoregressive time-series. Explicit forms for the power spectra and covariance functions of first- and second-order forms are given. A conceptually simple method is described for fitting continuous time autoregressive models to data. Formulae giving the standard errors of the parameter estimates are derived. Simulated data are used to verify the performance of the methods. Irregularly spaced observations of the two hydrogen-deficient stars FQ Aqr and NO Ser are analysed. In the case of FQ Aqr the best-fitting model is of second order, and describes a quasi-periodicity of about 20 d with an e-folding time of 3.7 d. The NO Ser data are best fitted by a first-order model with an e-folding time of 7.2 d.

Key words: methods: data analysis – methods: statistical.

1 INTRODUCTION

Time-series data such as those in Figs 1–2 (below) challenge the traditional methods of analysis used by astronomers. Typically, a frequency domain analysis of some sort would be attempted. Prominent peaks in a power spectrum would be selected, and the data 'explained' as being due to cyclical variations with superimposed noise. If there are no sufficiently dominant peaks in the spectrum, the conclusion may be that there is nothing much to be learned from the data.

Yet it is clear that there *is* information in the data: observations which are close together in time are generally more similar than those which are widely separated. Put differently, the data are not pure noise, but show correlation. Of course, the information is of a statistical, rather than a deterministic, nature: only probabilistic, rather than definite, predictions about future observations can be made.

In the case of observations which are regularly spaced in time, it is possible to progress beyond the mere investigation of the correlation structure of stochastic time-series. The reason is that *any* stationary stochastic time-series $x_t (t = 1, 2, \cdots, N)$ with mean μ can be written as

$$x_t = \mu + \sum_{j=1}^{p} \alpha_j [x_{t-j} - \mu] + \sum_{i=1}^{q} \beta_i \epsilon_{t-i} + \epsilon_t \,, \tag{1}$$

where the α_j and β_i are constants, and ϵ_k is white noise (e.g. Box & Jenkins 1976; Brockwell & Davis 1991). Equation (1) defines an 'autoregressive moving average' time-series of orders p and q, conveniently abbreviated ARMA(p,q). In practice quite modest values of p and q – typically $0 \leqslant p, q \leqslant 2$ – are usually enough to achieve adequate descriptions of stochastic time-series.

Series with $p = 0$, pure moving average or ARMA(0,q)≡MA(q) processes, are characterized by short coherence time-scales q. Long 'memory' processes, on the other hand, can often be described by pure autoregressive models of low order, i.e. ARMA(p,0)≡AR(p) forms with small p (typically $p \leqslant 2$). The subject of this paper is the modelling of time-series such as those in Figs 1 and 2 (below), i.e. observations with long coherence times. This means that our interest is in the AR(p) form

$$x_t = \mu + \sum_{j=1}^{p} \alpha_j [x_{t-j} - \mu] + \epsilon_t . \tag{2}$$

There is an enormous literature on the modelling of stochastic time-series observed at regularly spaced time points, and all major statistical software packages contain facilities for fitting standard models to such data. By contrast, work on the modelling of stochastic time-series observed at irregularly spaced time points is much less well known. Nonetheless, there is an extensive theoretical literature on the topic – see for example the papers in Parzen (1983). Although there are different approaches to dealing with data such as those in Fig. 1 (below), we will concentrate on the analogues of AR models mentioned above. In the economics literature these are often referred to as 'continuous time autoregressive' or 'CAR' models.

Whereas the AR(p) model (2) is a difference equation, CAR(p) models are defined by stochastic differential equations (SDEs), as is to be expected when generalizing to continuous time. The

relevant statistical theory of CAR(p) time-series is reviewed in the next section of this paper.

The next challenge is the fitting of CAR(p) models to observed time-series. A method suitable for small (say $N < 200$) data sets is described in Section 3. Evaluation of how well the model fits is also considered. Inference issues are dealt with in Section 4. The results of a few example analyses are presented in Section 5, and the paper is concluded in Section 6.

2 AN INTRODUCTION TO CAR(P) TIME SERIES

Much of the required theory is given by Jones (1981), and we adhere largely to his notation; see also Jones (1993) and Harvey (1989). A general form of the SDE satisfied by a CAR(p) process $x(t)$ is

$$\sum_{j=0}^{p} \alpha_j x^{(j)}(t) = \varepsilon(t)\,. \tag{3}$$

In this equation, the α_j are constants ($\alpha_p = 1$); $\varepsilon(t)$ is a continuous-time white noise process with variance σ_ε^2; and $x^{(j)}$ is the jth derivative of $x(t)$ [with $x^{(0)}(t) \equiv x(t)$]. As in the case of discrete-time processes, most CAR processes are of low order, typically $p \leqslant 2$.

The interpretation of (3) is not entirely straightforward, as the derivatives of an irregularly varying processes such as Brownian motion do not strictly speaking exist. The simplest form of the equation is

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) + \alpha_0 x(t) = \epsilon(t) \tag{4}$$

which can be written in differential form as

$$\mathrm{d}x(t) = -\alpha_0 x(t)\,\mathrm{d}t + \mathrm{d}w(t)\,. \tag{5}$$

In (5), $\mathrm{d}w$ is to be interpreted as the infinitesimal increment in a continuous time random walk (or Brownian motion) over the time increment $\mathrm{d}t$. Formally

$$\mathrm{d}w(t) = \epsilon(t)\,\mathrm{d}t\,,$$

i.e. $\epsilon(t)$ is the derivative of a random walk process. Readers interested in details of this, and other, formal theoretical aspects of SDEs are referred to Bergstrom (1984).

The second-order form of (3) can be written as

$$\begin{aligned} \mathrm{d}x^{(1)}(t) &= -\left[\alpha_0 x(t) + \alpha_1 x^{(1)}\right]\mathrm{d}t + \mathrm{d}w(t) \\ \mathrm{d}x(t) \quad &= x^{(1)}(t)\,\mathrm{d}t\,. \end{aligned} \tag{6}$$

A little reflection shows that solutions of (6) may, in general, be more smoothly varying (relatively more power at low frequencies) than those of (5): solutions of (5) are essentially integrations over $\mathrm{d}w(t)$, i.e. they will have roughly the character of random walks. Similarly, solutions for $x^{(1)}$ in (6) will be have the appearance of random walks, so that $x(t)$ will behave like an integrated random walk – i.e. be smoother than a random walk. We conclude that the higher the order p of the CAR(p) equation, the greater the degree of smoothness of the solutions it may have.

Some key results on the covariance functions and spectra of $x(t)$ satisfying (3) follow. The interested reader is referred to Doob (1953) for background material. The roots $r_j (j = 1, 2, \ldots, p)$ of the characteristic equation

$$A(z) = \sum_{j=0}^{p} \alpha_j z^j = 0 \tag{7}$$

are of some importance. As an example, the covariance function of $x(t)$ is given by

$$\begin{aligned} C(\tau) &= \mathrm{cov}[x(t), x(t+\tau)] \\ &= -\sigma_\varepsilon^2 \sum_{j=1}^{p} \frac{e^{r_j \tau}}{2\mathrm{Re}(r_j)\Pi_{k\neq j}(r_k - r_j)\left(r_k^* + r_j\right)} \end{aligned} \tag{8}$$

where $\mathrm{Re}(r_j)$ is the real part of r_j, and r_k^* is the complex conjugate of r_k.

The frequency spectrum of $x(t)$ is given by

$$S(\omega) = \frac{\alpha_0^2 C(0)}{\left|\sum_{j=0}^{p} \alpha_j (i\omega)^j\right|^2} \equiv \frac{\alpha_0^2 \sigma_x^2}{|A(i\omega)|^2}\,. \tag{9}$$

Since (7) can be written in terms of its roots as

$$A(z) = \Pi_{j=1}^{p}(z - r_j) = 0, \tag{10}$$

an equivalent form of (9) is

$$S(\omega) = \frac{\alpha_0^2 \sigma_x^2}{\Pi_{j=1}^{p}|r_j - i\omega|^2}\,. \tag{11}$$

In terms of the real and imaginary parts of the roots $r_j = a_j + ib_j$,

$$S(\omega) = \frac{\alpha_0^2 \sigma_x^2}{\Pi_{j=1}^{p}\left[a_j^2 + (\omega - b_j)^2\right]}\,. \tag{12}$$

For purely real roots r_j the spectrum $S(\omega)$ decreases monotonically with increasing frequency from a maximum at $\omega = 0$. However, for complex roots, local maxima occur at $\omega = b_j$: these correspond to quasi-periodicities in the data.

It is of interest to examine the explicit forms of (8) and (12) for the two lowest-order CAR(p) processes.

(i) If $p = 1$, then (7) reduces to

$$A(z) = \alpha_0 + z = 0 \quad \Rightarrow r_1 = -\alpha_0$$

and (8) to

$$C(\tau) = \frac{\sigma_\varepsilon^2}{2\alpha_0} e^{-\alpha_0 \tau}\,. \tag{13}$$

It follows that the variance and the autocorrelation function of $x(t)$ are

$$\sigma_x^2 = C(0) = \frac{\sigma_\varepsilon^2}{2\alpha_0} \qquad \rho(\tau) = C(\tau)/C(0) = e^{-\alpha_0 \tau}\,. \tag{14}$$

Clearly α_0^{-1} is a measure of the correlation memory of the process, and should be positive for physically realistic time-series. Note that for a given σ_ε, the variance of $x(t)$ increases with decreasing α_0, i.e. as the correlation memory increases.

An extreme case is the SDE

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \varepsilon(t)\,,$$

i.e. $\alpha_0 \to 0$; this is a random walk which has infinite correlation memory and infinite variance.

For general α_0 the frequency spectrum is

$$S(\omega) = \frac{\alpha_0^2 \sigma_x^2}{\alpha_0^2 + \omega^2}\,. \tag{15}$$

As $\alpha_0 \to \infty$, so $S(\omega) \to \sigma_x^2$, i.e. continuous time processes resembling discrete time white noise can be modelled as CAR(1) time-series with large α_0.

(ii) In the case $p = 2$ the roots of (7) are

$$r_1, r_2 = \frac{-\alpha_1 \pm \sqrt{\alpha_1^2 - 4\alpha_0}}{2}.$$

(a) If $\alpha_1^2 > 4\alpha_0$ both roots are real and

$$C(\tau) = -\frac{\sigma_\varepsilon^2}{2\left(r_2^2 - r_1^2\right)}\left[\frac{e^{r_1\tau}}{r_1} - \frac{e^{r_2\tau}}{r_2}\right] \tag{16}$$

and hence

$$\sigma_x^2 = \frac{-\sigma_\varepsilon^2}{2r_1r_2(r_1 + r_2)} = \frac{\sigma_\varepsilon^2}{2\alpha_0\alpha_1} \tag{17}$$

$$\rho(\tau) = \frac{1}{r_2 - r_1}\left[r_2e^{r_1\tau} - r_1e^{r_2\tau}\right]. \tag{18}$$

Equation (18) generalizes (14) to the case where there are two different time-scales in the time-series. The requirement that $r_1, r_2 < 0$ [deduced from (18)] translates into $\alpha_0, \alpha_1 > 0$.

By (12)

$$S(\omega) = \frac{\alpha_0^2\sigma_x^2}{\left(r_1^2 + \omega^2\right)\left(r_2^2 + \omega^2\right)}, \tag{19}$$

which is again a monotonically decreasing function of frequency.

(b) If $\alpha_1^2 < 4\alpha_0$, r_1 and r_2 must be complex conjugates:

$$r_1 = a + ib \quad r_2 = a - ib$$

with

$$a = -\frac{1}{2}\alpha_1 \quad b = \frac{1}{2}\sqrt{4\alpha_0 - \alpha_1^2}.$$

The covariance function reduces to

$$C(\tau) = \frac{-\sigma_\varepsilon^2 e^{a\tau}}{4ab(a^2 + b^2)^{1/2}}\cos(b\tau + \phi)$$

$$\phi = \tan^{-1}\frac{a}{b} = \sin^{-1}\left(\frac{-\alpha_1}{2\alpha_0^{1/2}}\right)$$

$$\sigma_x^2 = \frac{-\sigma_\varepsilon^2\cos\phi}{4ab(a^2 + b^2)^{1/2}} = \frac{-\sigma_\varepsilon^2}{4a(a^2 + b^2)} = \frac{\sigma_\varepsilon^2}{2\alpha_0\alpha_1}$$

$$\rho(\tau) = \frac{e^{a\tau}}{b}\sqrt{a^2 + b^2}\cos(b\tau + \phi). \tag{20}$$

The quantity b is positive by its definition; $a < 0$ for (e.g.) a physically meaningful variance. It is not difficult to see that $\alpha_0, \alpha_1 > 0$ should hold.

The spectrum is

$$S(\omega) = \frac{\alpha_0\sigma_x^2}{[a^2 + (\omega - b)^2][a^2 + (\omega + b)^2]} \tag{21}$$

i.e. $x(t)$ is quasi-periodic with angular frequency b.

3 FITTING MODELS TO SMALL DATA SETS

The key ingredient of the maximum likelihood model fitting technique discussed in this section is expression (8) for the covariance of $x(t)$. Let the *observed* values of the time-series be $\{X(t_1), X(t_2), \cdots, X(t_N)\}$, then

$$X(t_j) = \mu + x(t_j) + e(t_j) \tag{22}$$

where μ is the mean and $e(t_j)$ is a measurement error, with mean zero and variance σ_e^2. The covariances of the $X(t_j)$ are given by

$$\mathbf{C}_X(i, j) = \mathrm{cov}[X(t_i), X(t_j)]$$

$$= \mathbf{C}_x(i, j) + \sigma_e^2\mathbf{I} = \begin{cases} C(0) + \sigma_e^2 & i = j \\ C(|t_i - t_j|) & i \neq j \end{cases} \tag{23}$$

The Gaussian log likelihood $\mathcal{L}$ of the vector

$$\mathbf{X} = \begin{bmatrix} X(t_1) \\ X(t_2) \\ \cdots \\ X(t_N) \end{bmatrix} \tag{24}$$

is easily written down using the information above. The elements of the covariance matrix

$$\mathbf{\Sigma} = \mathrm{cov}(\mathbf{X}, \mathbf{X}')$$

follow immediately from (23) and (8), and

$$\mathcal{L} = -\frac{1}{2}\left[N\log 2\pi + \log|\mathbf{\Sigma}| + (\mathbf{X} - \mu\mathbf{1})'\mathbf{\Sigma}^{-1}(\mathbf{X} - \mu\mathbf{1})\right] \tag{25}$$

is the corresponding Gaussian log likelihood. [In (25), $\mathbf{1}$ is a column vector of N elements equal to unity].

Estimation of the unknown parameters $\alpha_0, \alpha_1, \cdots, \alpha_{p-1}, \sigma_e^2, \sigma_\epsilon^2, \mu$ has thus been reduced to the numerical problem of maximization of the likelihood function with respect to these unknowns. The computational burden can be reduced by obtaining explicit expressions for two of the unknowns. Set

$$\mathbf{\Sigma}_* = \mathbf{\Sigma}/\sigma_\epsilon^2 \tag{26}$$

and use

$$\frac{\partial\mathcal{L}}{\partial\sigma_\epsilon^2} = 0$$

to deduce

$$\widehat{\sigma}_\epsilon^2 = \frac{1}{N}(\mathbf{X} - \mu\mathbf{1})'\mathbf{\Sigma}_*^{-1}(\mathbf{X} - \mu\mathbf{1}) \tag{27}$$

Note that $\mathbf{\Sigma}_*$ now contains the unknowns $\alpha_0, \alpha_1, \cdots, \alpha_{p-1}$ and the ratio

$$R = \sigma_e^2/\sigma_\epsilon^2. \tag{28}$$

An explicit estimator for μ can also be derived:

$$\frac{\partial\mathcal{L}}{\partial\mu} = 0$$

and it follows that

$$\widehat{\mu} = \mathbf{1}'\mathbf{\Sigma}_*^{-1}\mathbf{X}/\mathbf{1}'\mathbf{\Sigma}_*^{-1}\mathbf{1}. \tag{29}$$

In practice, the solution then proceeds as follows.

(i) Guess values for $\alpha_0, \alpha_1, \cdots, \alpha_{p-1}, R$.
(ii) Calculate the corresponding value for $\mathbf{\Sigma}_*$.
(iii) An estimator for μ follows from (29).
(iv) Using $\widehat{\mu}$ from (iii), an estimator for σ_ϵ^2 follows from (27).
(v) Substitution of (27) into the expression (25) for $\mathcal{L}$ gives

$$\mathcal{L} = -\frac{1}{2}\left[N(1 + \log 2\pi\sigma_\epsilon^2) + \log|\mathbf{\Sigma}_*|\right]. \tag{30}$$

(vi) Iteration over (i)–(v) continues until $\mathcal{L}$ reaches a maximum.
(vii) By (28),

$$\widehat{\sigma}_e^2 = \widehat{R}\widehat{\sigma}_\epsilon^2.$$

Despite the manoeuvering in the preceding paragraphs, model fitting remains computationally expensive for quite modest sample sizes, due to the repeated inversion of the $N \times N$ matrix $\mathbf{\Sigma}_*$. A much faster algorithm which avoids the inversion (in fact even calculation of $\mathbf{\Sigma}_*$) will be described in the next paper of this series. The method of this paper is valuable, as it is very forgiving of poor starting guesses for the unknown parameter values. It is therefore very useful

for application to small subsets of data to find approximate solutions. It also has the virtue of conceptual simplicity.

The model fit is evaluated by studying the residuals. The latter are estimated by

$$\mathbf{z} = \mathbf{L}^{-1}(\mathbf{X} - \widehat{\mu}\mathbf{1}) \tag{31}$$

where $\mathbf{L}$ is a lower triangular matrix obtained from the Cholesky decomposition of $\mathbf{\Sigma}$:

$$\mathbf{\Sigma} = \mathbf{L}\mathbf{L}'$$

(see e.g. Koen 2000).

For time-series measured at regular intervals, fundamental tests for randomness of the residuals are based on their autocorrelation and partial autocorrelation functions. Finding a suitable form for these functions, and their statistical properties, are suitable topics for research. The spectrum as estimated from the periodogram can still be used to investigate whether there are excesses of power in some frequency intervals. There are some outstanding questions though: the choice of frequencies in which to evaluate the periodogram, and appropriate smoothing formulae, being two. Of course, simply plotting the residuals may also be instructive.

Further discussion of residual evaluation is postponed to a later paper. For present purposes the autocorrelations defined by

$$r_k = c_k/c_0 \quad c_k = \frac{1}{N}\sum_{j=1}^{N-k}[z(t_j) - \overline{z}][z(t_{j+k}) - \overline{z}] \quad k = 1, 2, \ldots \tag{32}$$

are plotted. This means that the correlations of successive values, values two observations apart, three observations apart, etc., are studied. Put another way, r_k is the correlation between observations for which the time *indices* differ by k, without any regard to the time *intervals* between the observations. If the residuals are indeed white noise, then they will be entirely free of correlation. This is true regardless of the time order, or even the time spacing, of the residuals. A test based on (32) can therefore be seen as a *partial* test for zero correlation – a 'minimum requirement' to be fulfilled by the residuals. Of course, tests based on equation (32) are particularly stringent, as at least some residuals with closely similar indices will be closely spaced in time, and hence more likely to be correlated if the residuals are not white noise. The usual approximation

$$\text{S.E.}(r_k) \approx \frac{1}{\sqrt{N}} \tag{33}$$

(e.g. Koen & Lombard 1993 and references therein) is used to gauge the significance of the autocorrelations.

Finally, the choice between different models that all fit satisfactorily but which may differ in complexity (e.g. number of fitted parameters n) can be based on a comparison between their information criteria,

$$\text{AIC} = -2\mathcal{L} + 2n \quad \text{BIC} = -2\mathcal{L} + n\log N. \tag{34}$$

AIC and BIC are, respectively, the Akaike and Bayes information criteria. In the case of the null model of (Gaussian) white noise $n = 2$ since the data can then be fully described by its mean and variance.

In keeping with results often quoted in the literature, it was found that the AIC consistently selects models with too many parameters. Therefore, only the BIC is used below.

It is a well-known statistical result that the maximum likelihood estimates of parameters ψ_j approach the true values as the sample size increases (e.g. Cox & Hinkley 1974). Furthermore, the covariance matrix of the large sample estimates of the ψ_j is given by

$$\mathcal{C} = \mathcal{I}^{-1} \tag{35}$$

where $\mathcal{I}$ is the Fisher information matrix defined by

$$\mathcal{I}_{ij} \equiv -\text{E}\left[\frac{\partial^2\mathcal{L}}{\partial\psi_i\partial\psi_j}\right].$$

The components of $\mathcal{I}$ are derived in Appendix A.

4 EXAMPLES

4.1 Simulated data

The simulated data in Figs 1 and 2 were calculated at 100 time points, spaced at exponentially distributed intervals. The values $\mu = 0$, $\sigma_\epsilon = 1$ were used throughout; $\sigma_e = 0.5$ in Fig. 1, and $\sigma_e = 2$ in Fig. 2. Table 1 contains the results of fitting both CAR(1)

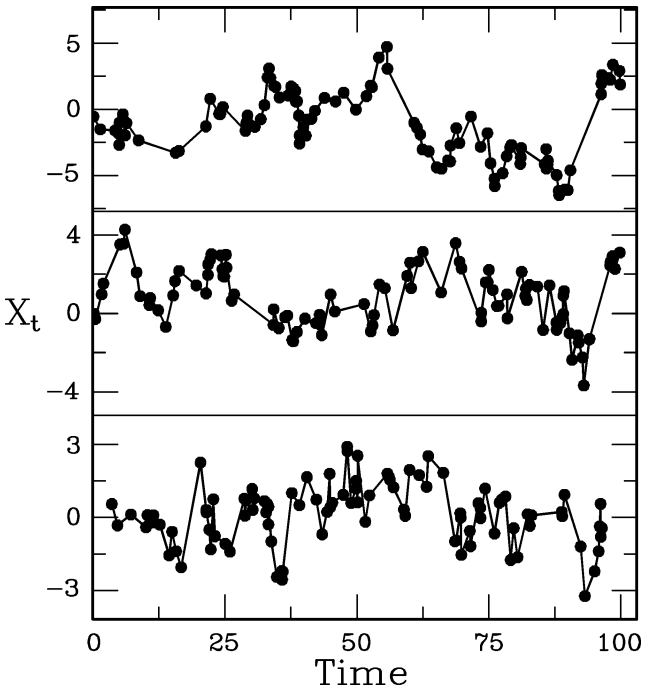

Figure 1. Simulated CAR(1) processes: from top to bottom, α_0 is 0.1, 0.2 and 0.5. In all three cases $\sigma_\epsilon = 1$, $\sigma_e = 0.5$.

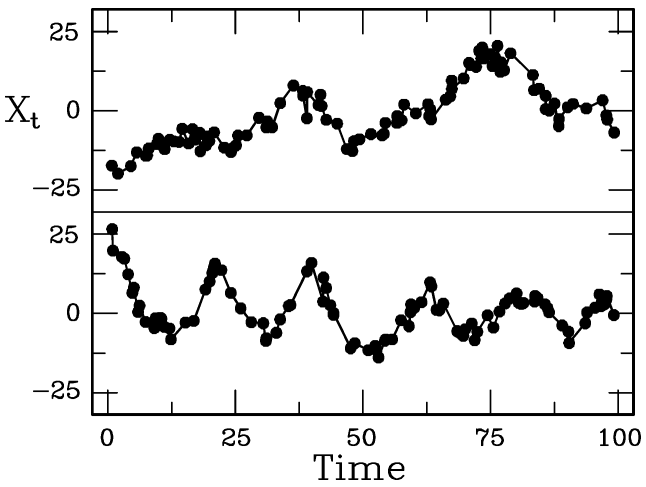

Figure 2. Simulated CAR(2) processes. In the top panel, $\alpha_0 = 0.01$, $\alpha_1 = 0.25$; in the bottom panel $\alpha_0 = 0.1$, $\alpha_1 = 0.1$. In both cases $\sigma_\epsilon = 1$, $\sigma_e = 2$.

Table 1. Estimation results for the simulated data of Figs 1 and 2. The first line of each block gives the true parameter values; the second the first-order solution, and the third the second-order solution. The asymptotic standard errors of estimates are given in brackets.

μ	σ_ϵ	σ_e	α_0	α_1	BIC
0.00	1.00	0.50	0.10	0.00	
−0.95 (0.9)	1.19 (0.2)	0.46 (0.09)	0.12 (0.06)		322.30
−1.02 (0.8)	2.77 (2.0)	0.46 (0.08)	0.34 (0.2)	2.05 (1.6)	324.65
0.00	1.00	0.50	0.20	0.00	
0.90 (0.4)	1.17 (0.2)	0.35 (0.08)	0.34 (0.1)		284.43
0.90 (0.3)	5.72 (5.4)	0.40 (0.07)	1.83 (1.7)	4.46 (4.5)	287.10
0.00	1.00	0.50	0.50	0.00	
0.11 (0.3)	1.08 (0.2)	0.52 (0.09)	0.45 (0.2)		292.71
0.11 (0.3)	6.18 (12)	0.56 (0.09)	2.66 (5.0)	5.67 (11)	296.86
0.00	1.00	2.00	0.01	0.25	
−5.48 (6.1)	2.40 (0.4)	2.01 (0.3)	0.031 (0.02)		545.09
−3.77 (4.0)	0.93 (0.3)	2.29 (0.2)	0.022 (0.01)	0.22 (0.1)	540.87
0.00	1.00	2.00	0.10	0.10	
2.60 (3.7)	3.64 (0.5)	1.49 (0.3)	0.089 (0.05)		560.99
0.58 (1.2)	1.21 (0.3)	2.00 (0.2)	0.10 (0.02)	0.10 (0.06)	296.86

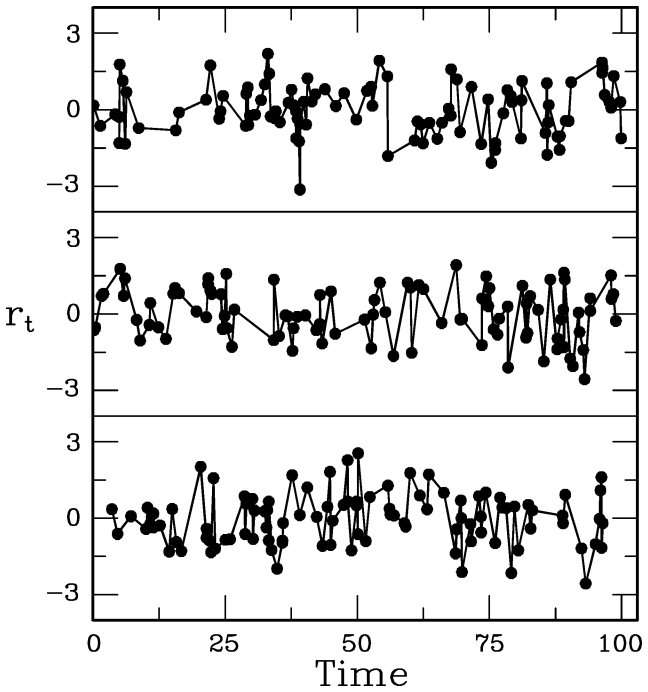

Figure 3. Residuals of the optimal models fitted to the three simulated data sets plotted in Fig. 1.

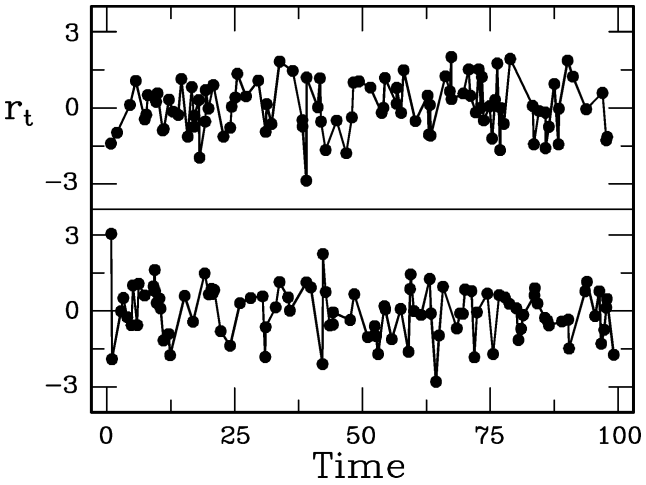

Figure 4. Residuals of the optimal models fitted to the two simulated data sets plotted in Fig. 2.

and CAR(2) models to each of the five data sets. The residuals calculated from (31) are plotted in Figs 3 and 4: clearly most, if not all, of the systematic parts of the variability have been removed. The autocorrelation functions defined in (32) are shown in Figs 5 and 6, for the 'best' (according to the BIC) models. Aside from a slight excess at lag $k = 1$ for the first data set, the correlation in the residuals appears negligible.

Inspection of Table 1 leads to the following conclusions.

(1) Provided the correct model is fitted, the parameter estimates are reasonably accurate. On the other hand, if the wrong model is fitted, the parameter estimates are very wide of the mark.

(2) The BIC indicates the correct model in all five cases.

(3) The values of the asymptotic standard errors appear to be quite reasonable.

4.2 Observations of FQ Aqr and NO Ser

Kilkenny et al. (1999) presented observations of the two hydrogen-deficient variable stars FQ Aqr (BD+1 4381) and NO Ser (BD−1 3438). For the former star $N = 177$ measurements spread over five observing seasons were obtained; for NO Ser, $N = 120$ over four seasons. The V-band data are plotted in Figs 7 and 8; clearly the time spacing is rather irregular, with large gaps between observing seasons.

Kilkenny et al. (1999) investigated previous claims in the literature, based on very little data, that the two stars show periodic variations. Based on their more extensive data Kilkenny et al. (1999) conclude, from the results of a Fourier analysis, that there are no coherent periodicities in the data.

The question addressed here is whether there is any evidence in the two data sets for stochastic cycles with characteristic frequencies,

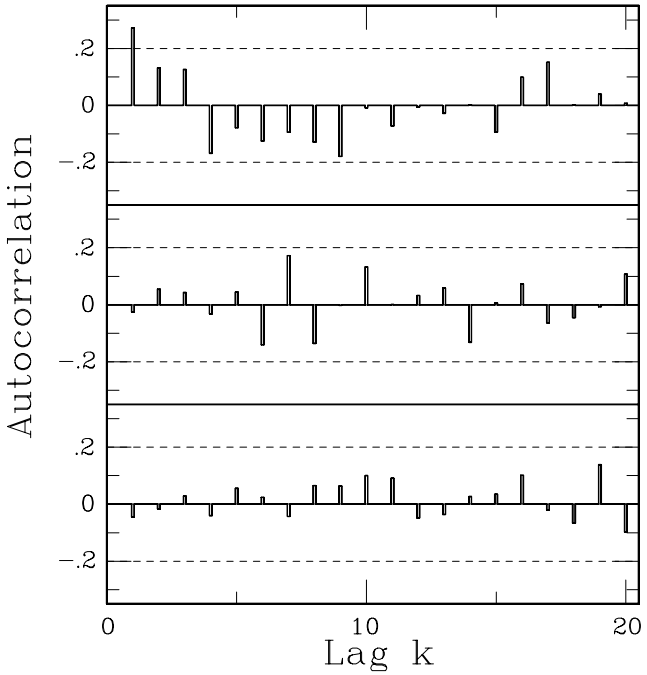

Figure 5. The autocorrelation functions of the residuals in Fig. 3.

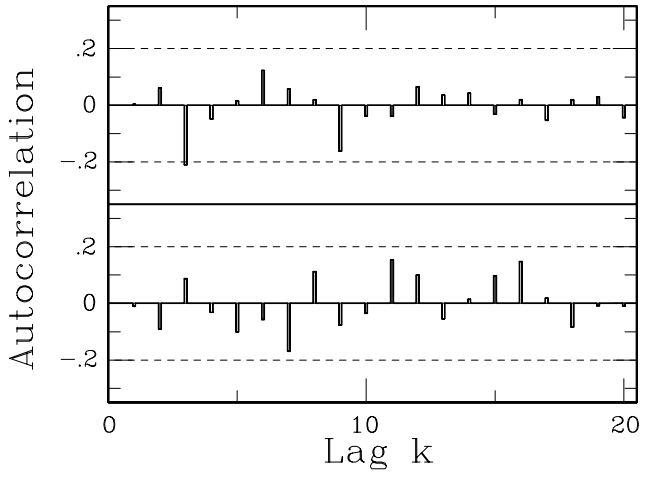

Figure 6. The autocorrelation functions of the residuals in Fig. 4.

i.e. periodicities which are not necessarily coherent. The results of fitting CAR(1) and CAR(2) models to the data are given in Table 2.

Interestingly, for both stars the estimated values of σ_e are zero in the CAR(1) case. The implication of this result is that the measurement error is negligibly small, and that sampling error is responsible for the zero result. Those models were therefore re-estimated assuming zero measurement error. The procedures given in Section 3, and in the Appendix, are affected very little by this assumption, as the reader will be able to verify readily. A complication arose in the CAR(2) model fit to the NO Ser data: the asymptotic covariance matrix C was ill-conditioned, and standard errors for the parameter estimated could therefore not be calculated. Fortunately, the difficulty is primarily an academic one, as the model is sub-optimal for the particular data set.

Inspection of Table 2 shows that the second-order model is selected for FQ Aqr. Referring to the discussion of CAR(2) models

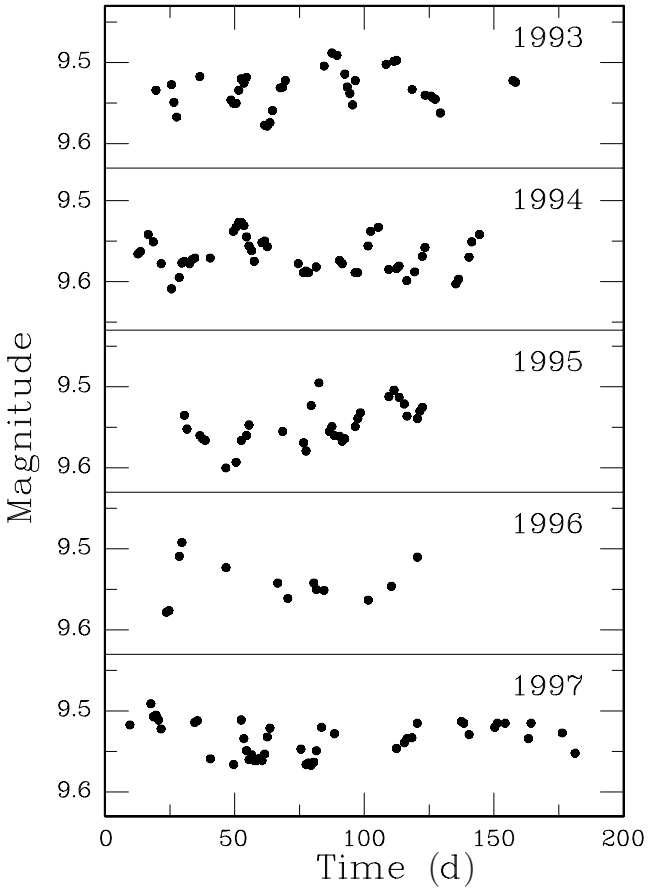

Figure 7. *V*-band observations over five seasons of the hydrogen-deficient variable FQ Aqr.

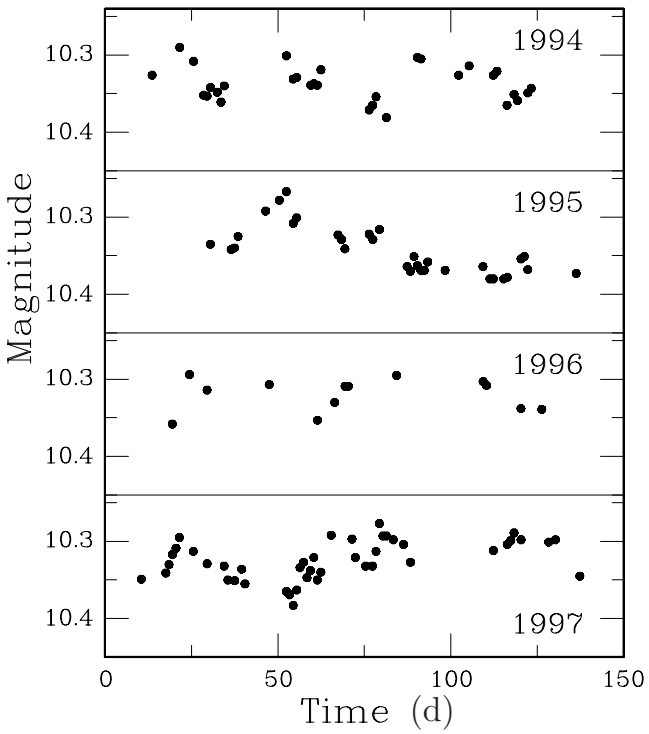

Figure 8. *V*-band observations over four seasons of the hydrogen-deficient variable NO Ser.

Table 2. Estimation results for the observations of FQ Aqr (first set of solutions) and NO Ser (second set of solutions). In both cases the CAR(1) estimates of the measurement error are effectively zero, so that second sets of solutions with the constraint $\sigma_e = 0$ were also calculated. Standard errors for the CAR(2) parameter estimates for NO Ser could not be calculated as the asymptotic covariance matrix $\mathcal{C}$ was ill-conditioned.

μ	σ_ϵ	σ_e	α_0	α_1	BIC
9.536 (3.5E-3)	0.012 (1.6E-3)	0	0.148 (4.0E-2)		−963.98
9.536 (3.5E-3)	0.012 (9.0E-4)	0	0.148 (3.7E-3)		−969.15
9.537 (2.7E-3)	0.010 (2.4E-3)	0.0039 (7.7E-4)	0.178 (3.9E-2)	0.54 (0.17)	−984.00
10.330 (4.8E-3)	0.014 (2.1E-3)	1.6E-5 (0.87)	0.14 (4.5E-2)		−612.57
10.330 (4.8E-3)	0.014 (1.2E-3)	0	0.139 (3.9E-3)		−617.36
10.330 (4.8E-3)	0.3864	0.0011	3.8484	27.3343	−608.18

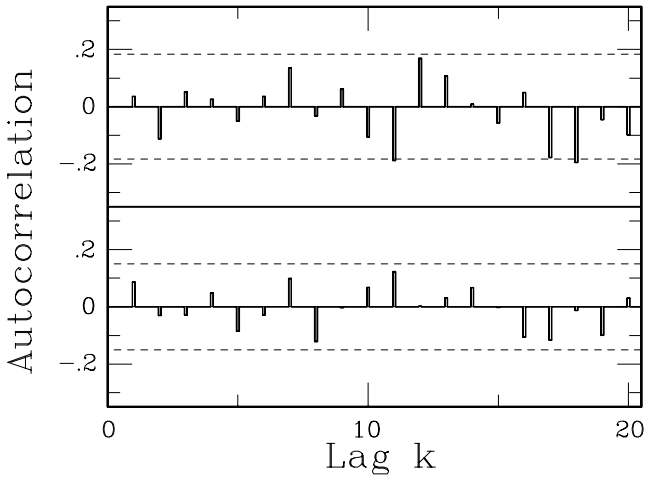

Figure 9. The autocorrelation functions of the residuals of the NO Ser data (top panel) and the FQ Aqr data (bottom panel).

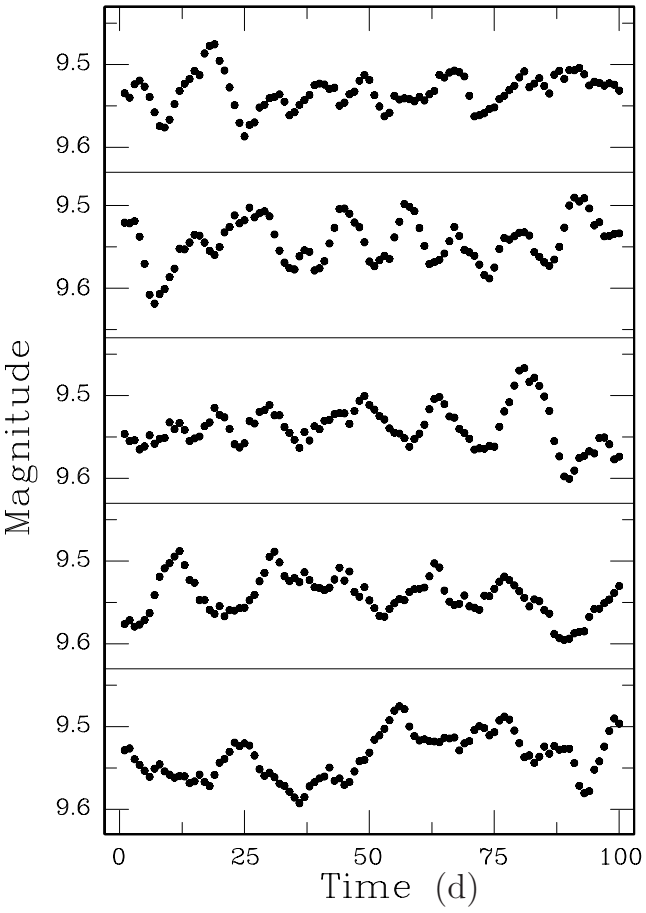

Figure 10. Five simulated data sets based on the parameters derived for FQ Aqr.

in Section 2, it can be seen that coefficients $\alpha_0 = 0.178$, $\alpha_1 = 0.54$ imply a quasi-periodicity with $b = 0.323$ ($P = 19.5$ d) and coherence time $1/a = 1/0.272 = 3.68$ d. Kilkenny et al. (1999) remarked 'Whether it is possible to infer a "characteristic" or "quasi"-period near 18 d, or whether the star exhibits essentially random variations, is difficult to say'. The present results endorse the former of the two conclusions, particularly as the BIC $= -823.75$ for the latter.

In the case of NO Ser, the CAR(1) model is preferred on the basis of its smaller BIC. Furthermore, the coefficients of the CAR(2) model have very large values – something which was seen in Table 1 when CAR(2) models were incorrectly fitted to CAR(1) data. The estimated coherence time-scale is $1/\alpha_0 = 7.2$ d.

The correlation in the residuals of the optimal models is negligible, at least as far as the test given in Section 4 is concerned – see Fig. 9. Figs 10 and 11 show a few simulated data sets based on the two optimal models in Table 2. It is interesting that periodic behaviour is sometimes very obvious in Fig. 10 (e.g. the second panel), but almost invisible at others (e.g. the last panel). Fig. 11 demonstrates instances in which 'periodicities' appear in data governed by a CAR(1) equation – see the third panel in particular.

5 CONCLUSIONS

Astronomers are very familiar with fitting deterministic continuous time models such as polynomials and sinusoids to discrete observations. This paper has addressed the issue of fitting stochastic continuous time models to irregularly spaced discrete observations. There are a number of potential benefits to be reaped from fitting SDEs to astronomical time-series.

(1) Rigorous estimates of variability time-scales, with known statistical properties such as uncertainties.

(2) Reliable estimates of the number of time-scales in the data (i.e. the order of the SDE), by use of information criteria and ensuring that model residuals are white noise.

(3) Testing for the presence of quasi-periodicities in cases where their presence may not be clear cut. [It is noted in passing that quasi-periodicities of various forms are seen in the light curves of a variety of astronomical objects, at a variety of time-scales. To give

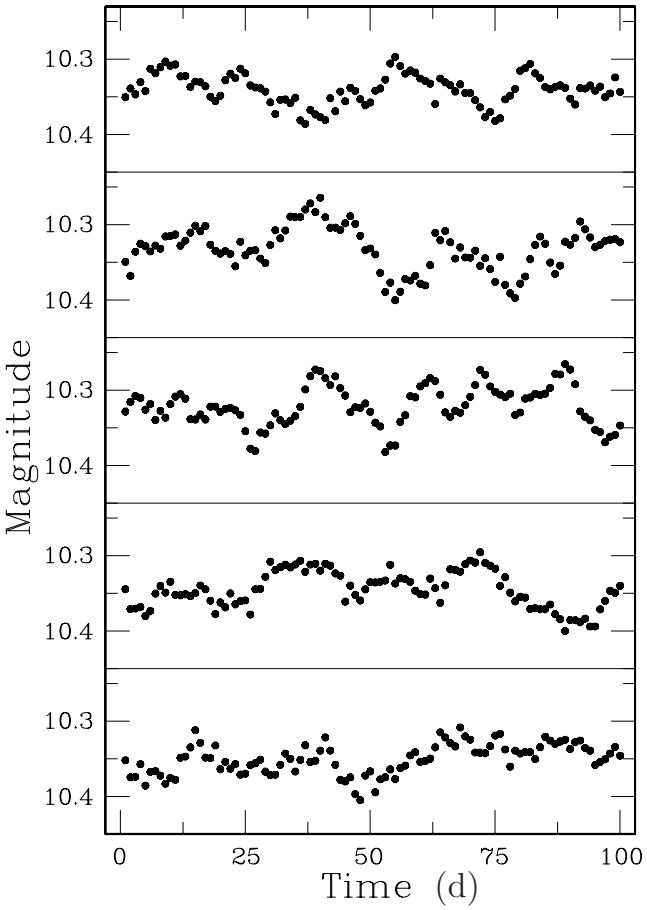

Figure 11. Five simulated data sets based on the parameters derived for NO Ser.

but two examples, quasi-periodicities in X-ray binaries may have quasi-periods of a few milliseconds (Van der Klis 2000) while in cataclysmic variables quasi-periods range from a few seconds to many minutes (Warner 2004)].

(4) Associated with the fitted SDE is a spectrum of the form seen in equation (9), i.e. the frequency spectrum of the data follow as a by-product of model fitting. The spectrum may lend itself to interpretation in terms of a physical model.

(5) Simulation of the variability of an object can only be performed if a reliable statistical model for the time-series of observations is available.

ACKNOWLEDGMENTS

The author is grateful to Dr Dave Kilkenny (SAAO) for making available the observations analysed in Section 5.

REFERENCES

Bargmann R. E., 1984, in Beyer W. H., ed., CRC Standard Mathematical Tables, 27th edn. CRC Press, Boca Raton, FL
Bergstrom A. R., 1984, in Griliches Z., Intriligator M. D., eds, Handbook of Econometrics, Vol. II. North-Holland, Amsterdam, p. 1146
Box G. E. P., Jenkins G. M., 1976, Time Series Analysis, Forecasting and Control. Holden-Day, Oakland, CA
Brockwell P. J., Davis R. A., 1991, Time Series: Theory and Methods. Springer-Verlag, New York
Cox D. R., Hinkley D. V., 1974, Theoretical Statistics. Chapman and Hall, London
Doob J. L., 1953, Stochastic Processes. John Wiley & Sons, New York
Harvey A. C., 1989, Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge Univ. Press, Cambridge
Jones R. H., 1981, in Findley D. F., ed., Applied Time Series Analysis II. Academic Press, New York
Jones R. H., 1993, Longitudinal Data with Serial Correlation: A State-Space Approach. Chapman & Hall, London
Kilkenny D., Lawson W. A., Marang F., Roberts G., Van Wyk F., 1999, MNRAS, 305, 103
Koen C., 2000, MNRAS, 316, 613
Koen C., Lombard F., 1993, MNRAS, 263, 287
Parzen E., ed., 1983, Lecture Notes in Statistics 25, Time Series Analysis of Irregularly Observed Data, Springer-Verlag, Berlin
Van der Klis M., 2000, ARA&A, 38, 717
Warner B., 2004, PASP, 116, 15

APPENDIX A: THE ASYMPTOTIC COVARIANCE MATRIX OF THE PARAMETER ESTIMATES

It is not difficult to show that the log likelihood (25) can also be written in the convenient form

$$\mathcal{L} = -\frac{1}{2}[N \log 2\pi + \log|\mathbf{\Sigma}| + \text{trace}(\mathbf{\Sigma}^{-1}\mathbf{G})] \tag{A1}$$

where

$$\mathbf{G} \equiv (\mathbf{X} - \mu\mathbf{1})(\mathbf{X} - \mu 1)' .$$

Using the rules of matrix differentiation (e.g. Bargmann 1984) it then follows that

$$\frac{\partial \mathcal{L}}{\partial \psi_j} = \text{trace}\left[(\mathbf{X} - \mu\mathbf{1})'\mathbf{\Sigma}^{-1}\mathbf{1}\frac{\partial \mu}{\partial \psi_j}\right] + \frac{1}{2}\text{trace}\left[\mathbf{\Sigma}^{-1}(\mathbf{G} - \mathbf{\Sigma})\mathbf{\Sigma}^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \psi_j}\right]$$

where the convenient notation

$$\boldsymbol{\psi} = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \cdots \\ \psi_{p+3} \end{bmatrix} = \begin{bmatrix} \mu \\ \sigma_\epsilon \\ \sigma_e \\ \alpha_0 \\ \alpha_1 \\ \cdots \\ \alpha_{p-1} \end{bmatrix} \tag{A2}$$

has been introduced. The Hessian of the log likelihood is

$$\frac{\partial^2 \mathcal{L}}{\partial\psi_i \partial\psi_j} = -\text{trace}\left[\frac{\partial\mu}{\partial\psi_i}\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}\frac{\partial\mu}{\partial\psi_j}\right] - \frac{1}{2}\text{trace}\left[\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_j}\right]$$
$$+ \text{trace}\left[(\mathbf{X}-\mu\mathbf{1})'\boldsymbol{\Sigma}^{-1}\left(\mathbf{1}\frac{\partial^2\mu}{\partial\psi_i\partial\psi_j} - \frac{\partial\boldsymbol{\Sigma}}{\partial\psi_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\mu}{\partial\psi_j}\mathbf{1} - \frac{\partial\boldsymbol{\Sigma}}{\partial\psi_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\mu}{\partial\psi_i}\mathbf{1}\right)\right]$$
$$+ \frac{1}{2}\text{trace}\left[\boldsymbol{\Sigma}^{-1}(\mathbf{G}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\left(\frac{\partial^2\boldsymbol{\Sigma}}{\partial\psi_i\partial\psi_j} - 2\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_j}\right)\right]$$

Taking expectations then gives the elements

$$\begin{aligned}\mathcal{I}_{ij} &\equiv -\text{E}\left[\frac{\partial^2\mathcal{L}}{\partial\psi_i\partial\psi_j}\right]\\ &= \text{trace}\left[\frac{\partial\mu}{\partial\psi_i}\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}\frac{\partial\mu}{\partial\psi_j}\right] + \frac{1}{2}\text{trace}\left[\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_j}\right]\\ &= \left[\frac{\partial\mu}{\partial\psi_i}\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}\frac{\partial\mu}{\partial\psi_j}\right] + \frac{1}{2}\text{trace}\left[\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_j}\right] \qquad \text{(A3)}\end{aligned}$$

of the Fisher information matrix. A similar result can be found in Harvey (1989).

It follows immediately from (A3) that

$$\mathcal{I}_{11} = \mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}$$

$$\mathcal{I}_{1j} = \mathcal{I}_{j1} = 0 \quad j \neq 1\,.$$

From (23) and (8),

$$\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_2} = \frac{2}{\sigma_\epsilon}\mathbf{C}_x$$

$$\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_3} = 2\sigma_e\mathbf{I}\,.$$

Also needed are terms of the form

$$\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_j} = \frac{\partial\mathbf{C}_x}{\partial\alpha_{j-4}} \quad j = 4, 5, \cdots, p+3\,.$$

For a first order SDE,

$$\frac{\partial C(\tau)}{\partial\alpha_0} = -\frac{\sigma_\epsilon^2}{2\alpha_0^2}e^{-\alpha_0\tau}(1+\alpha_0\tau) = -C(\tau)(1+\alpha_0\tau)$$

from (13). For the second order, case differentiation of (16) gives

$$\begin{aligned}\frac{\partial C(\tau)}{\partial\alpha_k} &= \frac{\sigma_\epsilon^2}{2(r_2^2-r_1^2)^2}\left\{2\left(\frac{e^{r_2\tau}}{r_2} - \frac{e^{r_1\tau}}{r_1}\right)\left(r_2\frac{\partial r_2}{\partial\alpha_k} - r_1\frac{\partial r_1}{\partial\alpha_k}\right)\right.\\ &\qquad \left. -e^{r_2\tau}(1-r_2\tau)\left[1-\left(\frac{r_1}{r_2}\right)^2\right]\frac{\partial r_2}{\partial\alpha_k} - e^{r_1\tau}(1-r_1\tau)\left[1-\left(\frac{r_2}{r_1}\right)^2\right]\frac{\partial r_1}{\partial\alpha_k}\right\}\\ &= 2C(\tau)\left(r_2\frac{\partial r_2}{\partial\alpha_k} - r_1\frac{\partial r_1}{\partial\alpha_k}\right)\\ &\qquad + \frac{\sigma_\epsilon^2}{2\left(r_2^2-r_1^2\right)^2}\left[\frac{e^{r_1\tau}}{r_1^2}(1-r_1\tau)\frac{\partial r_1}{\partial\alpha_k} - \frac{e^{r_2\tau}}{r_2^2}(1-r_2\tau)\frac{\partial r_2}{\partial\alpha_k}\right]\end{aligned}$$

and it is readily shown that

$$\frac{\partial r_1}{\partial\alpha_0} = \frac{-1}{\sqrt{\alpha_1^2-4\alpha_0}} \qquad \frac{\partial r_2}{\partial\alpha_0} = \frac{1}{\sqrt{\alpha_1^2-4\alpha_0}}$$

$$\frac{\partial r_1}{\partial\alpha_1} = \frac{-r_1}{\sqrt{\alpha_1^2-4\alpha_0}} \qquad \frac{\partial r_2}{\partial\alpha_1} = \frac{r_2}{\sqrt{\alpha_1^2-4\alpha_0}}\,.$$

In general (8) can be differentiated to find

$$\frac{\partial C}{\partial\alpha_i} = -\frac{\sigma_\epsilon^2}{2}\sum_{j=1}^{p}\frac{\exp(r_j\tau)}{\text{Re}(r_j)F_j}\left\{\tau\frac{\partial r_j}{\partial\alpha_i} - \text{Re}\left(\frac{\partial r_j}{\partial\alpha_i}\right)\Big/\text{Re}(r_j) \; -\sum_{\ell nej}^{p}\left[\frac{1}{r_\ell - r_j}\left(\frac{\partial r_\ell}{\partial\alpha_i} - \frac{\partial r_j}{\partial\alpha_i}\right) + \frac{1}{r_\ell^* + r_j}\left(\frac{\partial r_\ell^*}{\partial\alpha_i} + \frac{\partial r_j}{\partial\alpha_i}\right)\right]\right\}$$

where

$F_j \equiv \Pi_{k\neq j}(r_k - r_j)(r_k^* + r_j)$.

The partial derivatives of the roots r_k can be found by differentiating the expression

$$\sum_{j=0}^{p} \alpha_j r_k^j = 0$$

with respect to α_i to find

$$r_k^i + \sum_{j=0}^{p} j\alpha_j r_k^{j-1} \frac{\partial r_k}{\partial \alpha_i} = 0.$$

The last equation gives

$$\frac{\partial r_k}{\partial \alpha_i} = -r_k^i \Bigg/ \sum_{j=0}^{p} j\alpha_j r_k^{j-1} \quad .$$

This paper has been typeset from a TEX/LATEX file prepared by the author.